Cointegration and Vector Error Correction Representation

Katalin Varga (Ph.D.)

BUTE

2016 Autumn

Definition of Cointegration

- Suppose y_t and w_t are each integrated, ie. have unit roots:

$$(1-L)y_t = a(L)\delta_t$$
$$(1-L)w_t = b(L)v_t$$

If a linear combination exists which is stationary:

$$y_t - \alpha w_t$$

y_t and w_t are said to be cointegrated: [1 -α] is their cointegrating vector.

Some plausible examples:

1. Log GDP and log consumption
2. Log GDP and log investment
3. Money and price level

Interesting Facts about Cointegrated Variables

- If y_t and w_t cointegrated, [1 -α] is their cointegrating vector, then their regression is superconsistent:

$$y_t = \beta w_t + u_t$$

So $\beta \rightarrow \alpha$ even if the noise u_t is correlated with w_t, and at a faster rate than usual.

When we are investigating consumption functions, we have to deal with the solution of simultanious equations:

$$y_t = c_t + a_t$$
$$c_t = \alpha y_t + \varepsilon_t$$

We have unbiased, consistent estimate of α, if

$$a_t = a_{t-1} + \delta_t$$

So y and thus c have unit root property, but $c_t - \alpha y_t = \varepsilon_t$ is stacionary.

Representation of Cointegrated Systems

- Let x_t be a difference stacionary vector, if there exist a scalar vector α, so that $\alpha^T x_t$ is stacionary, then the vector x_t is cointegrated.
- Since the differences of x_t are stationary, x_t has a MA representation:

$$(1 - L)x_t = A(L)\varepsilon_t$$

So the fact that $\alpha^T x_t$ is stacionary implies extra restrictions on $A(L)$, and it must involve a restriction on $A(1)$.

- Still valid the multivariate Beverage- Nelson decomposition:

Then $y_t = c_t + z_t$

Where $z_t = \mu + z_{t-1} + A(1)\varepsilon_t$

And $c_t = A^*(L)\varepsilon_t$; $A_j{}^* = -\sum_{k=j+1}^{\infty} A_k$.

The Rank Condition on A(1)

- Multivariate Beveridge–Nelson decomposition:

$$x_t = z_t + c_t$$
$$(1 - L)z_t = A(1)\varepsilon_t$$
$$c_t = A^*(L)\varepsilon_t; \;\; A^*{}_j = -\sum_{k=j+1}^{\infty} A_k$$

- The elements of x_t are cointegrated with α_i iff ${\alpha_i}^T A(1) = 0$. This implies that the rank of $A(1)$ is (number of elements of x_t - number of cointegrating vectors α_i)

1. Case: $A(1) = 0 \leftrightarrow x_t$ stacionary in levels; all linear combinations of x_t stationary in levels.

2. Case: $A(1)$ $less\ than\ full\ rank \leftrightarrow (1 - L)x_t$ stacionary, some linear combinations $\alpha^T x_t$ stacionary

3. Case: $A(1)$ $full\ rank \leftrightarrow (1 - L)x_t$ stacionary, no linear combinations of x_t stacionary

Common Trend Representation

$\Psi = A(1)\Sigma A(1)^T$, the covariance of the innovation of the random walk component in the B-N decomposition.

When the rank of this matrix is deficient, we need fewer than N random walk components to describe the N series. This means that there are **common random walk** components.

- In the case of a two-dimensional cointegrated system the random walk components are perfectly correlated:

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} z_t + stacionary\ component$$

$$\Psi = A(1)\Sigma A(1)^T = Q\Lambda Q^T (symmetric)$$

If the system has N series and K cointegrating vectors, the rank of Ψ is N-K, K of the eigenvalues are zero. Let ν_t be a new error sequence:

$$\nu_t = Q^T A(1)\varepsilon_t$$

Common Trend Representation

- So $E(\nu_t {\nu_t}^T) = \Lambda$, is diagonal.
- In terms of the new shocks the B-N trend is:

$$z_t = z_{t-1} + A(1)\epsilon_{t-1} = z_{t-1} + Q\nu_t$$

- Since z_t and z_t are perfectly correlated we can write the system with only one random walk:

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} z_t + A^*(L)\varepsilon_t$$

$$(1 - L)z_t = \nu_{1t} = \begin{bmatrix} 1 & 0 \end{bmatrix} Q^T A(1)\varepsilon_t$$

This is the common trend representation, y_t and w_t share a single common trend or common random walk component.

The Impulse Response Function

- If we calculate the impulse response of x_t in level, then after cumulating it, we get A(1): that is $A(1)_{yw}$ is the long term response of y to a unit shock in w.
- Let $\alpha = [1 - 1]^T$, so that $\alpha^T A(1) = 0$.

$$[1 - 1]\begin{bmatrix} A(1)_{yy} & A(1)_{yw} \\ A(1)_{wy} & A(1)_{ww} \end{bmatrix} = 0$$

So the long term response of the variables:

$$A(1)_{yy} = A(1)_{wy}$$
$$A(1)_{yw} = A(1)_{ww}$$

Error Correction Representation

- In autoregressive representation:

$$x_t = -B_1 x_{t-1} - B_2 x_{t-2} + \cdots + \varepsilon_t$$

- Applying the Beveridge-Nelson decomposition:

$$x_t = -(B_1 + B_2 + \cdots)x_{t-1} + \sum_{j=1}^{\infty} B^*{}_j \Delta x_{t-j} + \varepsilon_t$$

$$\Delta x_t = -B(1)x_{t-1} + \sum_{j=1}^{\infty} B^*{}_j \Delta x_{t-j} + \varepsilon_t$$

Here the matrix $B(1)$ is responsible for the cointegration properties.

Since Δx_t, $\sum_{j=1}^{\infty} B^*{}_j \Delta x_{t-j}$ and ε_t are stationary, thus $B(1)x_{t-1}$ is stationary as well.

Error Correction Representation

- There are fundamentally three cases:

Case 1.

$B(1)$ is of full rank, then every linear combination of x_{t-1} and x_{t-1} itself is stacionary: we **should** run the VAR in **levels**.

Case 2.

The rank of $B(1)$ is strictly between 0 and full rank. Some linear combinations are stacionary, so x_t is cointegrated. If we run the VAR in levels, it will be **consistent**, but it **won't be efficient**.

We **must not** run the VAR in differences, it will be **misspecified**.

Case 3.

The rank of $B(1)$ is 0 , then none of the linear combinations of x_{t-1} is stacionary. Δx_t is stacionary without any cointegrations: we **should** run the VAR in first differences.

Error Correction Representation

- If $B(1)$ has less than full rank, then $B(1) = \gamma\alpha^T$.

- If there are k cointegrating vectors, the the rank of $B(1)$ is k.

- Then both γ and α have k coloumns, rewriting the system with γ and α:

$$\Delta x_t = -\gamma\alpha^T x_{t-1} + \sum_{j=1}^{\infty} B^*{}_j \Delta x_{t-j} + \varepsilon_t$$

- Here α is the matrix of cointegrating vectors. Since $\alpha^T x_{t-1}$ is stationary $-\gamma\alpha^T x_{t-1}$ is stationary as well.
- $\gamma\alpha^T x_{t-1}$ is correcting the errors occuring in Δx_{t-j} and shifting $\alpha^T x_t$ towards its mean, so it has the „meanreverting” property.

Cointegration with Drifts and Trends

- Suppose we put back the trend μ in the equation:

$$(1-L)x_t = \mu + A(L)\varepsilon_t$$

The B-N decomposition is then:

$$z_t = \mu + z_{t-1} + A(1)\varepsilon_t$$

We have two choices:

1. $\alpha^T x_t$ stationary and $\alpha^T \mu = 0$ (this is a separate restriction)

2. $\alpha^T A(1) = 0$ but $\alpha^T \mu \neq 0$, then:

$$\alpha^T z_t = \alpha^T \mu + \alpha^T z_{t-1} \rightarrow \alpha^T z_t = (\alpha^T \mu)\mathrm{t} + \alpha^T z_0$$

Thus $\alpha^T x_t$ will contain a time trend plus a stationary component.

Alternatively we can define cointegration to be $\alpha^T x_t$ contains a time trend but no stochastic trends.

The Johansen Tests

- The Johansen Tests are likelihood ratio tests.

- Let's take the VECM form of the x_t:

$$\Delta x_t = -B(1)x_{t-1} + \sum_{j=1}^{\infty} B^*{}_j \Delta x_{t-j} + \varepsilon_t$$

$$B(1) = -\left(I - \sum_{j=1}^{\infty} B_j\right)$$

The rank of $B(1)$ is equal to the number of independent cointegrating vectors.

Maximum Eigenvalue Test

Maximum Eigenvalue Test:

The first test is a test whether the rank of the matrix is zero. The null hypothesis is that rank (B(1)) = 0 and the alternative hypothesis is that rank (B(1)) = 1. For further tests, the null hypothesis is that rank (B(1)) = 1; 2;… and the alternative hypothesis is that rank (B(1)) = 2; 3; …

$$\lambda_{max}(r, r+1) = -T ln\left(1 - \hat{\lambda}_{r+1}\right) \quad \text{maximum eigenvalue test}$$

This likelihood ratio statistic does not have the usual asymptotic 2 distribution. This is similar to the situation for the Dickey-Fuller test: the unit roots in the data generate nonstandard asymptotic distributions.

Trace Test

Trace Test:

The trace test is a test whether the rank of the matrix is r_0. The null hypothesis is that rank (B(1)) = r_0.

The alternative hypothesis is that r_0 < rank(B(1)) $\leq n$, where n is the maximum number of possible cointegrating vectors.

For the succeeding test if this null hypothesis is rejected, the next null hypothesis is that rank (B(1)) = r_0 + 1 and the alternative hypothesis is that r_0 + 1 < rank (B(1)) $\leq n$.

Testing proceeds as for the maximum eigenvalue test.

The likelihood ratio test statistic is:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln\left(1 - \hat{\lambda}_i\right) \quad \text{trace test}$$

Trace Test

Why is the trace test called the „trace test"?

It is called the trace test because the test statistic's asymptotic distribution is the trace of a matrix based on functions of Brownian motion or standard Wiener processes (Johansen Econometrica 1995, p. 1555).

The test is not based on the trace of B(1). But it is informative on the rank of B(1).

Literature

1. Lutkepohl, H. (2005). New introduction to multiple time series analysis. *Econometric theory*, *22*(5), 961-967.
2. Johansen, S. (1995). Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. Journal of econometrics, 69(1), 111-132.
3. Campbell, J. Y., & Perron, P. (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots. In NBER Macroeconomics Annual 1991, Volume 6 (pp. 141-220). MIT press.